Data-Driven Excellence: Navigating the Landscape of Football Player Performance

A Business Intelligence Report Analysing The 2022/2023 Premier League Season.

COUSIN, BARCLAY (Student)

9/1/24

BSc Information Technology

Contents

1.0 Executive Summary

Upon conducting a thorough analysis of the 2022/2023 Premier League statistics, several key findings have emerged allowing for a valuable insight into the individual and team performance and trends within data.

One key finding relates to the Premier League itself, the average age of players within the division is 25 years meaning the Premier League is maintaining its stature and is not considered a “retirement league” nor “development league” The steadfast maintenance of a 25-year-old average reaffirms the premier leagues status as a highly competitive league where players are typically in the prime year of their career. The physicality of the league remains to an athletic standard with yellow cards averaging 1.8 cards per game, per team showcasing a remarkable level of physical exertion combined with a very high average distance covered per game with the highest total achieved by Manchester City with 104 thousand metres accumulated over 38 games.

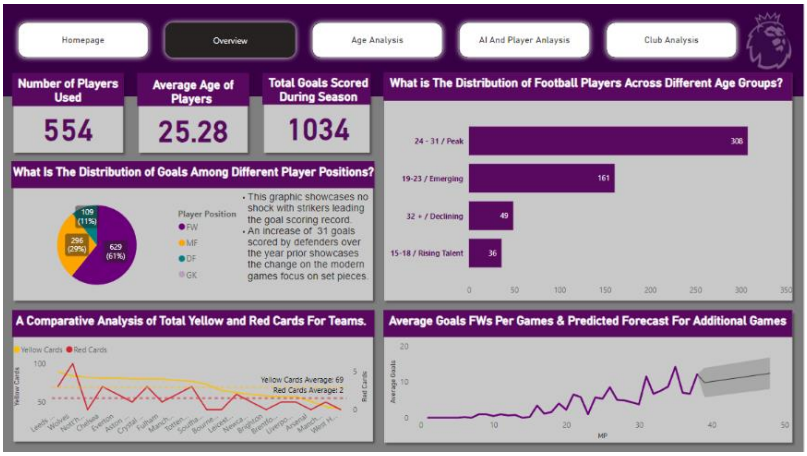


Figure 1: Overview Page

Figure 2: Age Analysis Page

An investigation was executed into what can be considered the prime age for a footballer. Some findings show that 22-year-olds score the most goals but as Erling Haaland broke the goal record aged 22 scoring 36 goals this result is skewered. After analysis, it was found that 24, 25, and 26-year-olds got the most premier league minutes last season and 22, 25, 26-year-old strikers scored the most goals. Interestingly the most total distance covered was between 23 to 26-year-olds and a decline occurred after 28 yet a surprising peak occurred at 31 years of age. A player is dispossessed the least between the ages of 27 and 29. This information can provide good evidence that a football player's peak years are between the ages of 24 and 29 so scouting teams should understand this.

Player analysis was conducted, and some stand-out over performing and underperforming players were outlined with a specific focus on finding players that were outside the “big 6” clubs. It is no surprise that on the best-performing midfielders graphic, 4 of 7 players obtained a summer transfer move. A recommendation is made that Moises Caicedo and Declan Rice are potential signings for the club and will be a good fit as a box-to-box midfielder. As all 3 players currently play in England they will be a good fit into the social core due to their understanding of the culture and language. Some defenders which are recommended to look at further are Kenny Tete, Fabian Schar, and Antonee Robinson due to their strong performances last season. The star standout player is James Ward-Prowse with him ranking 5th for outperforming his xG and strong midfielder statistics.

Figure 3: AI And Player Analysis Page

2.0 Introduction to The PowerBI Project

2.1 Introduction

The English Premier League is arguably the largest footballing league in the world with billions of pounds both invested and generated as profit. However, data is not easily accessible for clubs and data analysts are required to be hired to analyse this data. This Project aims to offer an insight into the 2022/2023 Premier League season to view factors that would be of use to football clubs as well as interest to football fans.

The dataset was chosen due to the close universities' connections with Middlesbrough Football club to help showcase what students can produce within the university and to see if they are impressed with some of the visuals. Features will be focused to compliment a range of data analysis from traditional grouping of data to the implementation of AI charts. This dataset addressed the Big Data problem as it there is a large volume, variety and velocity within the set.

The below questions were chosen to answer via a BI Dashboard.

Q1 - What Are Some Key Insights That Can Be Taken From The 2022/2023 Premier League Season?
Q2 - What insights can we gather from analysing player behaviour in the latest season? Can we identify the typical age range when players reach the peak of their footballing career? Are there clear patterns or trends that reveal strengths and weaknesses in different age categories, providing insights into optimal performance periods and potential areas for improvement?
Q3 - How can we identify potential hidden gems outside the top 6 clubs based on data analysis? Additionally, how can we leverage AI to enhance our understanding of player data, uncovering valuable insights that may contribute to scouting efforts and the discovery of untapped talent?
Q4 -What trends and factors can be discerned from club analysis, specifically exploring the reasons behind relegation for certain clubs and the factors contributing to Manchester City winning the league? Additionally, is there any observable relationship between squad diversity and on-field success, shedding light on whether diverse squads correlate with positive outcomes in league standings?

Figure 4: Questions Answered in The BI Dashboard

2.2 Data Source

2.2.1 Source

The chosen data set is the Premier League Player Statistics 2022-23 from Kaggle. Six individual files are included are required to be merged and transferred into a star schema.

2.2.2 Data Dictionary

Column	Description	Column	Description
nationID	Unique identifier for the nation.	Nation	Nation to which the player belongs.

Figure 5: Nation Table

Column	Description	Column	Description
PositionID	Unique identifier for the player's position.	Position	Specific position of the player.
PrimaryPosition	Player's primary playing position.	SecondaryPosition	Player's secondary playing position, if applicable.

Figure 6: Position Table

Column	Description	Column	Description
SquadID	Unique identifier for the team/squad.	Squad	Name or identifier for the team/squad the player plays for.

Figure 7: Squad Table

Column	Description	Column	Description
Player	Player's name.	Age	Age of the player.
Born	Birthdate or birth year of the player.	SquadID	Unique identifier for the team/squad.
NationID	Unique identifier for the nationality/country.	PlayerID	Unique identifier for the player.
PositionID	Unique identifier for the player's position.	AgeGroup	Age group category to which the player belongs.

Figure 8: Player Table

Column	Description	Column	Description
PlayerID	Unique identifier for the player.	Player	Player's name.
MP	Total number of appearances.	Starts	Number of starts.
Min	Total minutes played.	90s	Minutes per match (90 minutes equivalent).
Gls	Total number of goals scored.	Ast	Total number of assists.
G+A	Total goals and assists combined.	npGls	Goals excluding penalty kicks.
PK	Total penalty kicks scored.	PKatt	Total penalty kicks attempted.
CrdY	Number of yellow cards received.	CrdR	Number of red cards received.
xG	Expected goals based on shot quality.	npxG	Expected goals excluding penalties.
xAG	Expected assists based on shot quality.	npxG+xAG	Combined non-penalty xG and xA.
PrgC	Number of carries that move the ball forward.	PrgP	Number of passes that move the ball forward.
PrgR	Number of progressive passes received.	Glsp90	Goals per 90 minutes.
Astp90	Assists per 90 minutes.	G+Ap90	Combined goals and assists per 90 minutes.
npGlsp90	Non-penalty goals per 90 minutes.	npGlsAstp90	Combined non-penalty goals and assists per 90 mins.
xGp90	Expected goals per 90 minutes.	xAGp90	Expected assists per 90 minutes.
xG+xAGp90	Combined xG and xA per 90 minutes.	npxGp90	Non-penalty xG per 90 minutes.
npxG+xAGp90	Combined non-penalty xG and xA per 90 minutes.	Sh	Total number of shots taken.
SoT	Total number of shots on target.	SoT%	Percentage of shots on target.
Sh/90	Average shots per 90 minutes.	SoT/90	Average shots on target per 90 minutes.
G/Sh	Goals per shot ratio.	G/SoT	Goals per shot on target ratio.
Dist	Average distance of shots taken.	FK	Number of shots from free kicks.
npxG/Sh	Non-penalty xG per shot.	G-xG	Goals minus xG.
np:G-xG	Non-penalty goals minus xG.	Touches	Total number of touches.
Def Pen	Number of touches in the defensive penalty area.	Def 3rd	Number of touches in the defensive third.
Mid 3rd	Number of touches in the middle third.	Att 3rd	Number of touches in the attacking third.
Att Pen	Number of touches in the attacking penalty area.	Live	Number of live ball touches.
Att	Total number of attempted passes.	Succ	Total number of successful passes.
Succ%	Percentage of successful passes.	Tkld	Total number of tackles attempted.
Tkld%	Percentage of successful tackles.	Carries	Total number of ball carries.
TotDist	Total distance covered with the ball (Metres).	PrgDist	Total distance covered with the ball progressively.
1/3	Number of entries into the final third.	CPA	Completed passes into the penalty area.
Mis	Number of miscontrols.	Dis	Number of times dispossessed.
Rec	Number of ball recoveries.	Cmp	Number of completed passes.
Cmp%	Pass completion percentage.	Cmp_1	Completed passes in Zone 1.
Att_2	Pass attempts in Zone 2.	Cmp%_3	Pass completion percentage in Zone 3.
Cmp_4	Completed passes in Zone 4.	Att_5	Pass attempts in Zone 5.

Cmp%_6	Pass completion percentage in Zone 6.	Cmp_7	Completed passes in Zone 7.
Att_8	Pass attempts in Zone 8.	Cmp%_9	Pass completion percentage in Zone 9.
xA	Expected assists.	A-xAG	Difference between actual assists and xA.
KP	Key passes.	PPA	Passes leading to a shot.
CrsPA	Successful crosses leading to a shot.	Tkl	Total number of tackles.
TklW	Number of tackles won.	Tkl_1	Tackles in the defensive third.
Att.2	Total number of tackles attempted in the defensive third.	Tkl%	Tackle success percentage.
Lost	Number of times the player lost possession.	Blocks	Number of blocked shots.
Pass	Number of passes intercepted.	Int	Number of interceptions.
Tkl+Int	Total number of tackles and interceptions.	Clr	Number of clearances.
Err	Number of errors leading to a shot or goal.		

Figure 9: StatsData Table

3.0 Data Pre-Processing and Data Cleansing

Generally, the data itself was clean with limited Null values but many pre-processing steps were required to ensure data visualisation could occur in combination with a star schema.

3.1 Loading Data

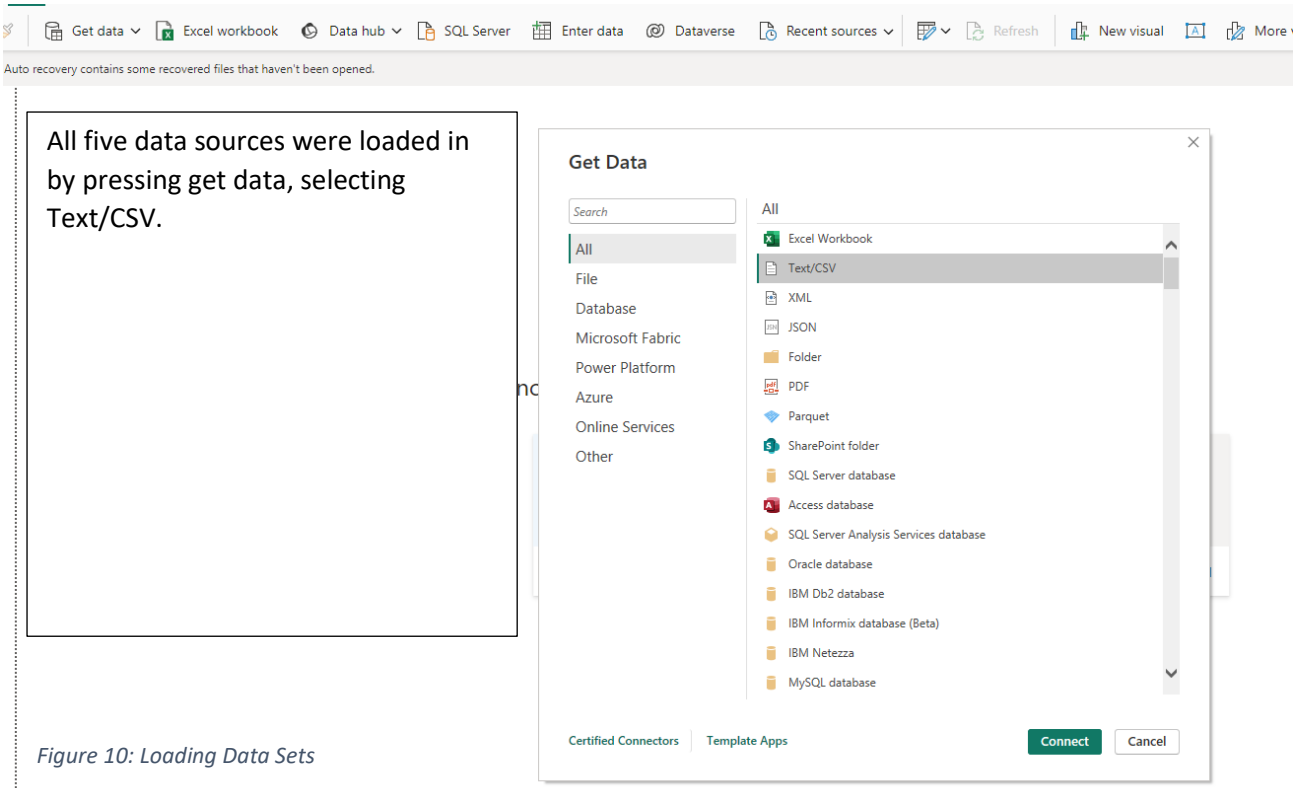


Figure 10: Loading Data Sets

3.2 Setting Column Headers

Before:

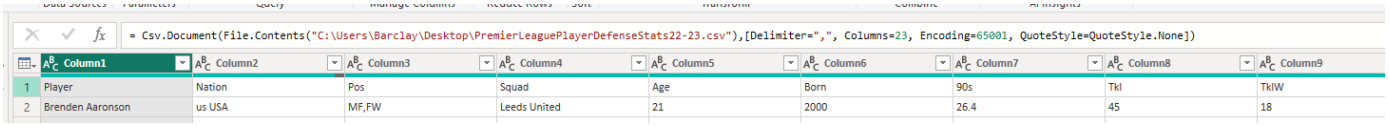


Figure 11: Unclean Imported Table

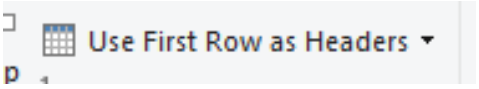


Figure 12: Use First Row as Headers Function

After:

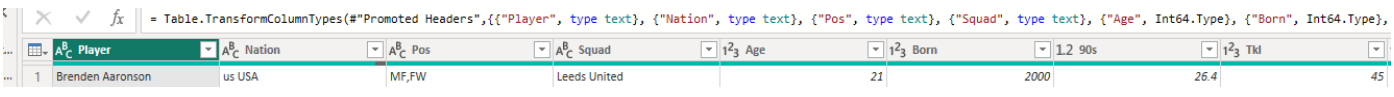
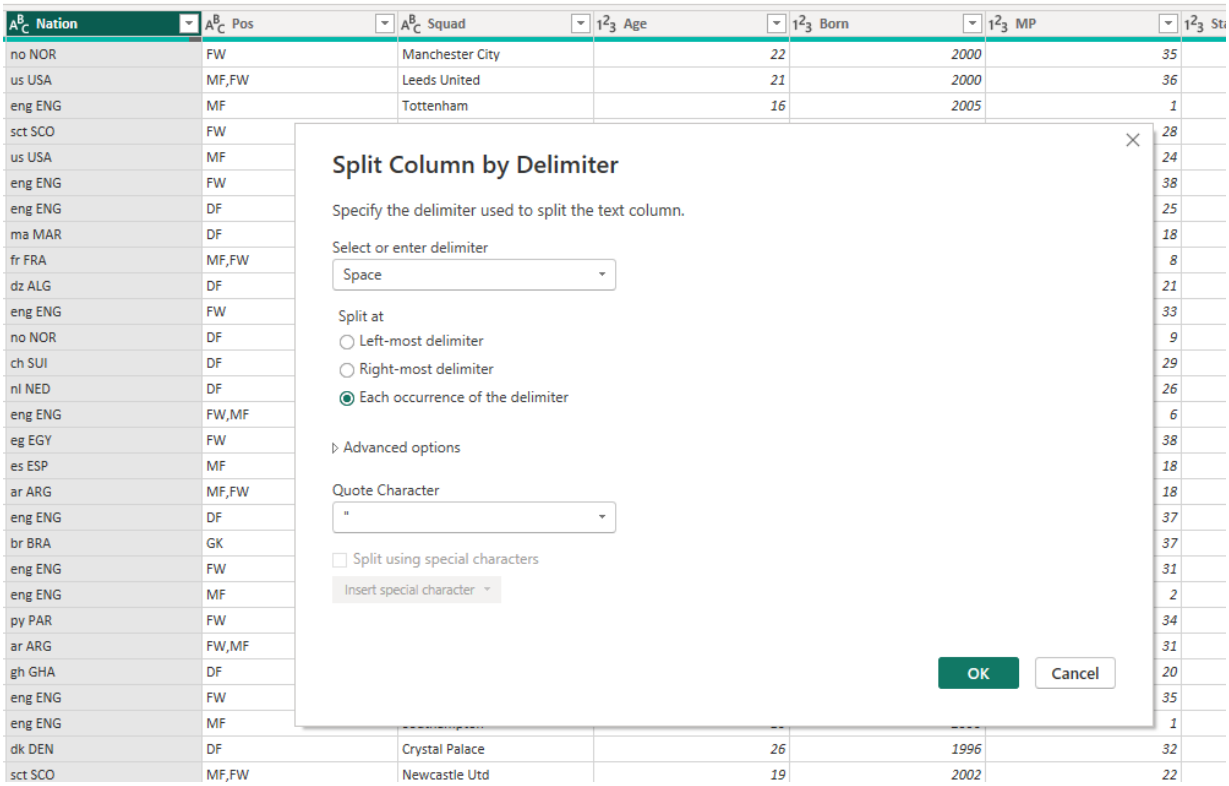


Figure 13: Clean Data With Correct Column Headers

3.3 Splitting Nationality Column

This was required as PowerBI Struggles to plot countries based on two letter values. The two-letter value table was deleted, and the 3-letter column was used.



Nation	Pos	Squad	Age	Born	MP	Sta
no NOR	FW	Manchester City	22	2000	35	
us USA	MF,FW	Leeds United	21	2000	36	
eng ENG	MF	Tottenham	16	2005	1	
sct SCO	FW				28	
us USA	MF				24	
eng ENG	FW				38	
eng ENG	DF				25	
ma MAR	DF				18	
fr FRA	MF,FW				8	
dz ALG	DF				21	
eng ENG	FW				33	
no NOR	DF				9	
ch SUI	DF				29	
nl NED	DF				26	
eng ENG	FW,MF				6	
eg EGY	FW				38	
es ESP	MF				18	
ar ARG	MF,FW				18	
eng ENG	DF				37	
br BRA	GK				37	
eng ENG	FW				31	
eng ENG	MF				2	
py PAR	FW				34	
ar ARG	FW,MF				31	
gh GHA	DF				20	
eng ENG	FW				35	
eng ENG	MF				1	
dk DEN	DF	Crystal Palace	26	1996	32	
sct SCO	MF,FW	Newcastle Utd	19	2002	22	

Figure 14: Splitting Nationality Column

The Pre-existing table was then deleted and replaced with the 3-letter abbreviation Table. Splitting Was also later used with the positions table as shown below.

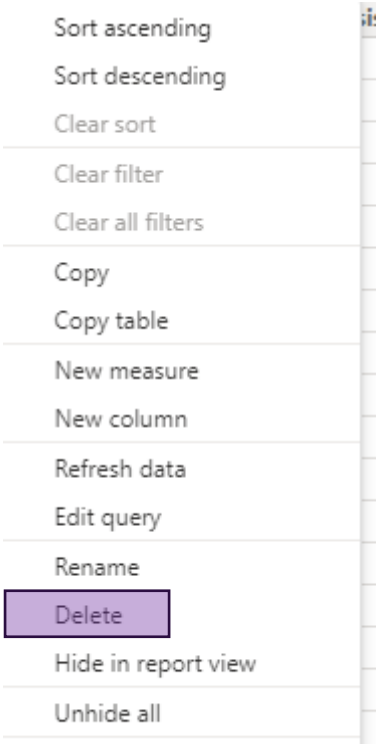


Figure 16: Delete Tables That Are Not Needed

	PositionID	Position	PrimaryPosition	SecondaryPosition
1	1	FW	FW	
2	2	DF	DF	
3	3	MF	MF	
4	4	MF,FW	MF	FW
5	5	FW,MF	FW	MF
6	6	GK	GK	
7	7	DF,MF	DF	MF
8	8	FW,DF	FW	DF
9	9	MF,DF	MF	DF
10	10	DF,FW	DF	FW

Figure 15: New Position Table Due to Splitting

3.4 Removing Nulls

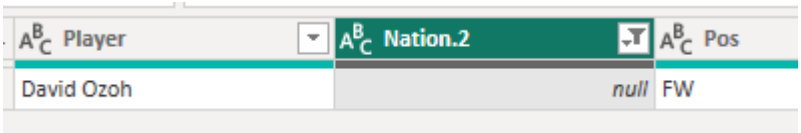


Figure 17: Player Without Nationality

The only null in the player database was for one player's nationality. As this was only one result, research was carried out and showed he was of Spanish nationality but was capped for England youth team, so England was the chosen nationality.

Before:

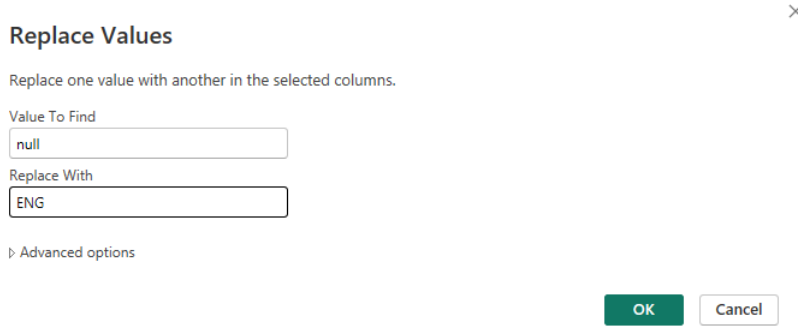


Figure 18 :Using Replace Values For Player Nationality

Moving Onto the PlayerStats Table, much of the table was already clean but it was found that any column that had a pre calculated percentage, had Nulls.

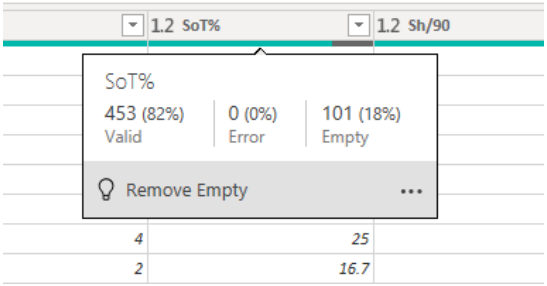


Figure 19: Identified Null Values In % Table

Null values were simply replaced with 0 for all the Null percentage categories as it showcases there is not enough data to calculate a percentage. This was executed on all other tables that contains a % figure.

After:

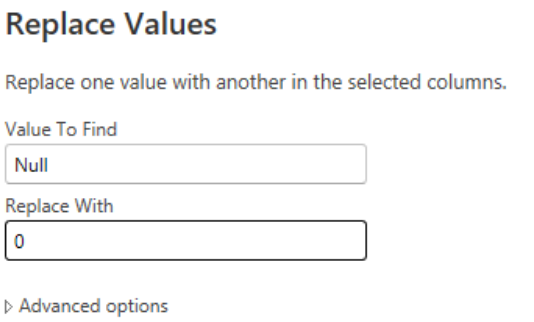


Figure 21: Replacing Null

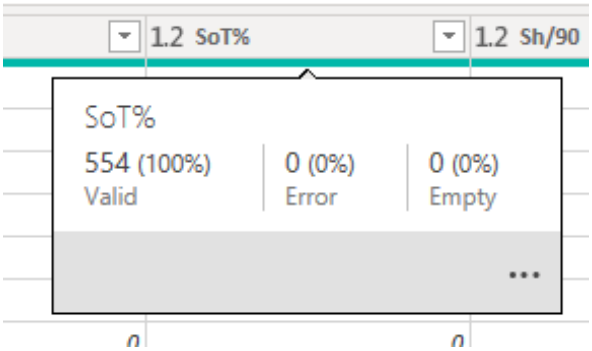


Figure 20: Percent Table with No Nulls

3.5 Combining All the Tables

1: Open Power Query Editor

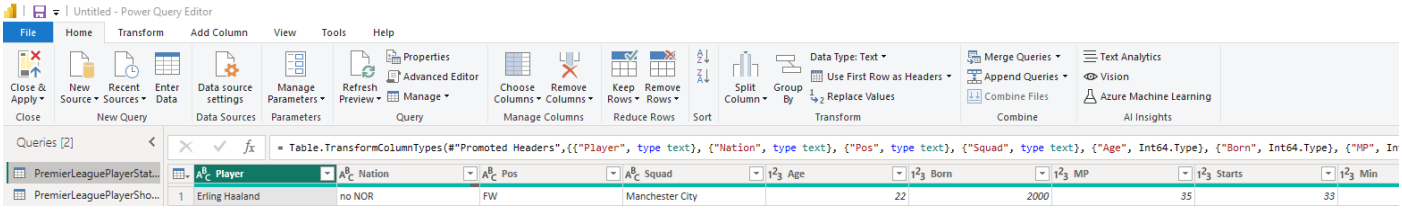


Figure 22: Power Query Editor

2: Merge Queries

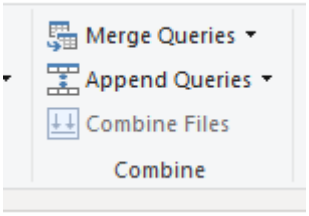


Figure 23: Merge Queries Option

3: Choose The Table to Merge and Select The Merging Column (Player)

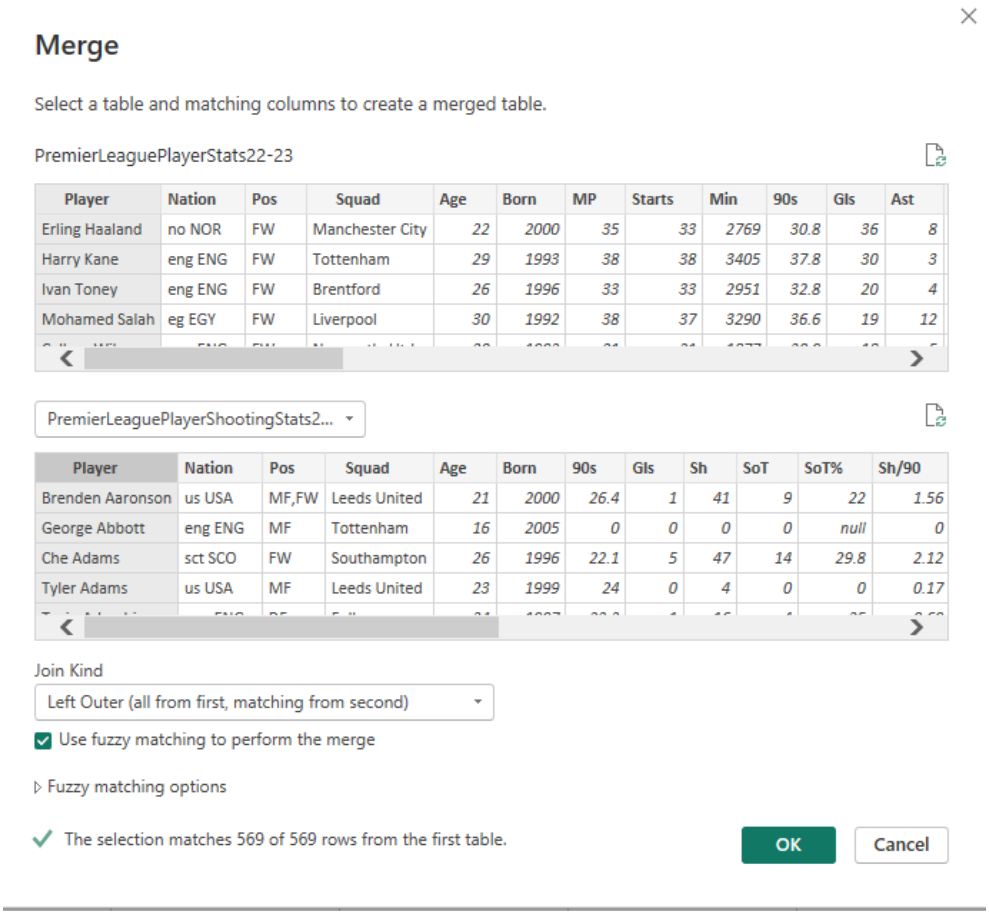


Figure 24: Merge Table Option with Player Selected

4: View The New Table

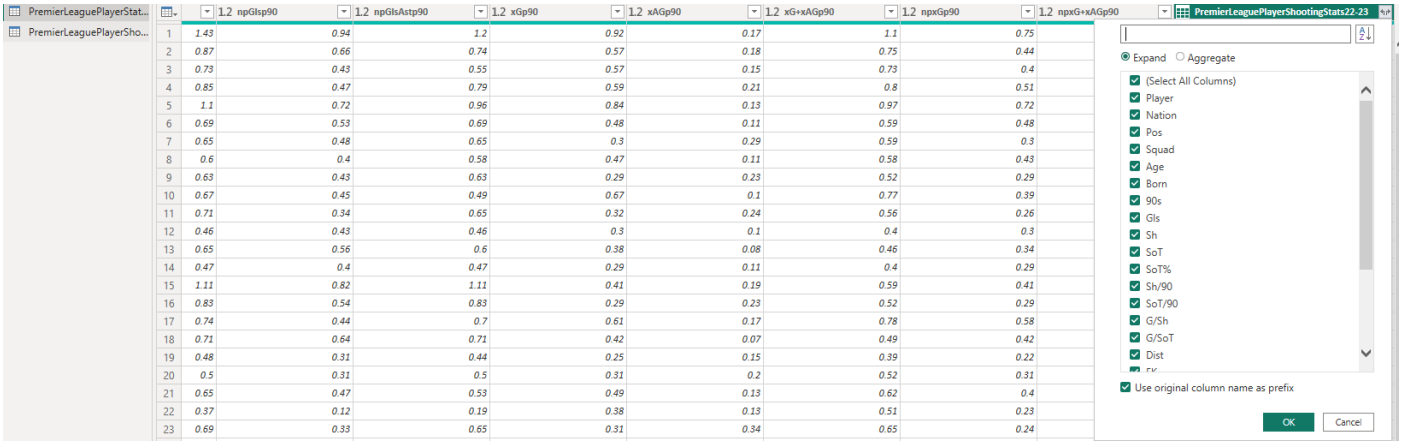


Figure 25: New Column Created Due to Merge

5: Select All Required Tables and Deselect Use Original Column Name As Prefix and Click Okay

	Player	Nation	Pos	Squad	Age	Born	MP	Starts	Min
1	Erling Haaland	no NOR	FW	Manchester City	22	2000	35	33	2
2	Harry Kane	eng ENG	FW	Tottenham	29	1993	38	38	3
3	Ivan Toney	eng ENG	FW	Brentford	26	1996	33	33	2
4	Mohamed Salah	eg EGY	FW	Liverpool	30	1992	38	37	3
5	Callum Wilson	eng ENG	FW	Newcastle Utd	30	1992	31	21	1
6	Marcus Rashford	eng ENG	FW	Manchester Utd	24	1997	35	32	2
7	Martinelli	br BRA	FW	Arsenal	21	2001	36	34	2
8	Ollie Watkins	eng ENG	FW	Aston Villa	26	1995	37	36	3
9	Martin Ødegaard	no NOR	MF	Arsenal	23	1998	37	37	3
10	Aleksandar Mitrović	rs SRB	FW	Fulham	27	1994	24	23	2
11	Bukayo Saka	eng ENG	FW	Arsenal	20	2001	38	37	3
12	Harvey Barnes	eng ENG	FW,MF	Leicester City	24	1997	34	32	2
13	Rodrigo	es ESP	FW,MF	Leeds United	31	1991	31	23	1
14	Rodrigo	es ESP	FW,MF	Leeds United	31	1991	31	23	1
15	Miguel Almirón	py PAR	FW	Newcastle Utd	28	1994	34	29	2
16	Roberto Firmino	br BRA	FW,MF	Liverpool	30	1991	25	13	1

Figure 26: New Table Created with Correct Data Extracted

6 View Successful Merge

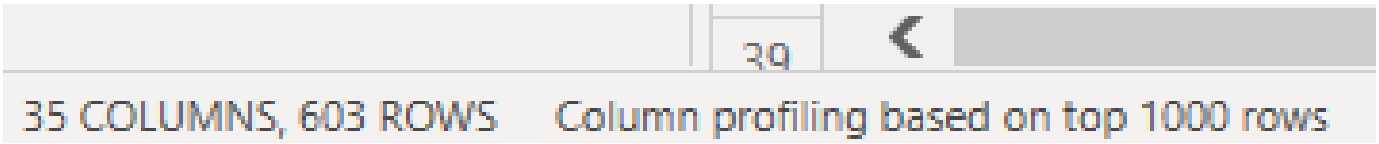


Figure 27: Proof of Successful Merge

7: Repeat for 4 Other Tables to Create Central table.

3.6 Creating the Star Schema

Creating the star schema requires separating one master table into 5 separate tables. This is done by the following steps.

3.6.1 Open In Power Query Editor

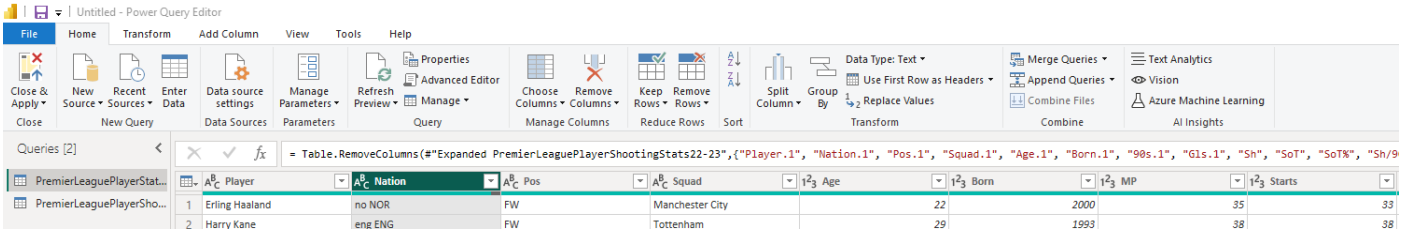


Figure 28: Power Query Editor

3.6.2 Right Click the Master Table, Duplicate the Table

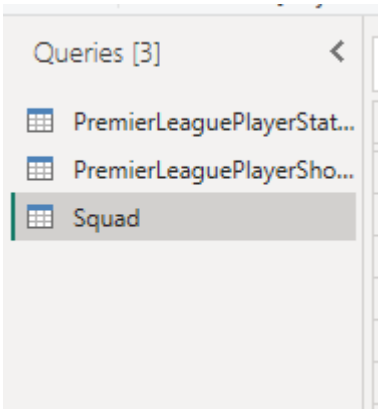


Figure 29: Table Duplication

3.6.3 Remove Unwanted Columns

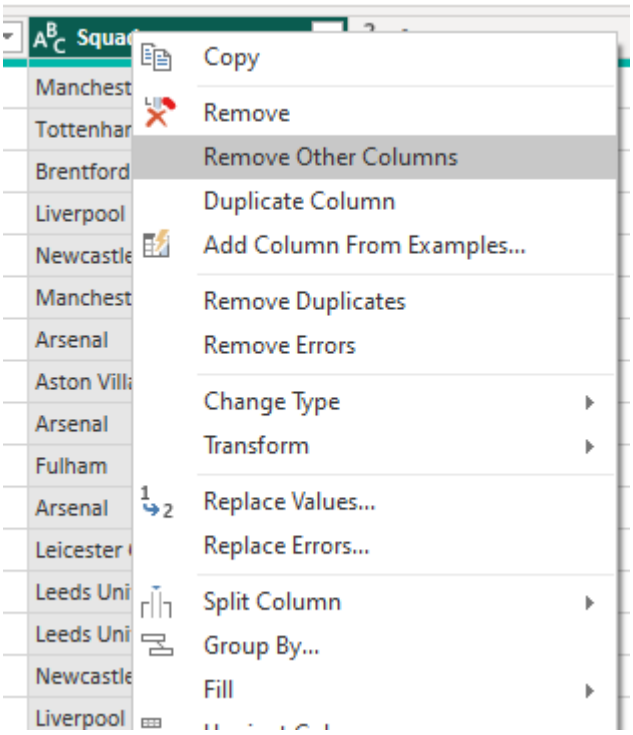


Figure 30: Removing Any Duplicate Columns

3.6.4 Remove Duplicates

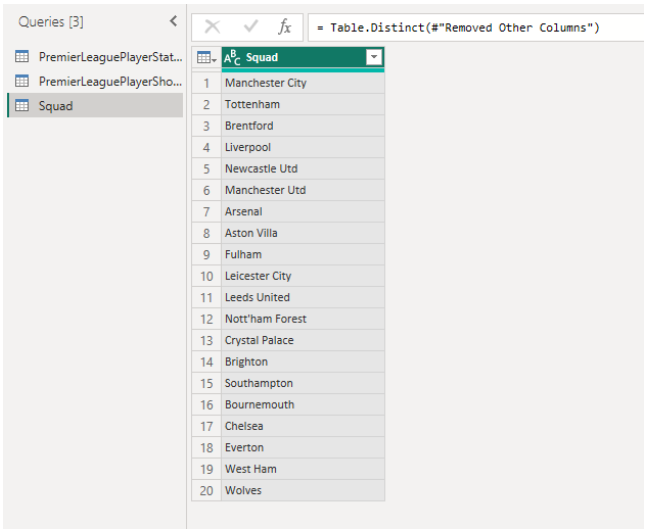


Figure 31: New Table with No Duplicates

3.6.5 Create Index Column

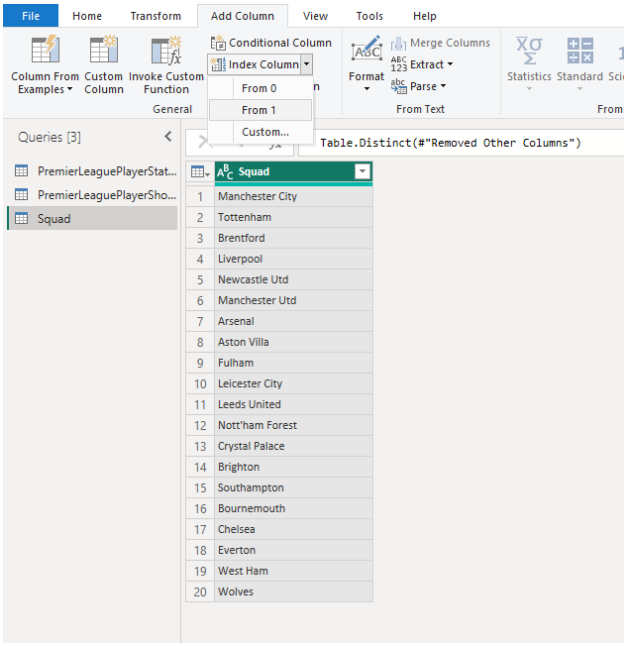


Figure 32: Creation of Index Column

3.6.6 Merge Queries of Primary Table

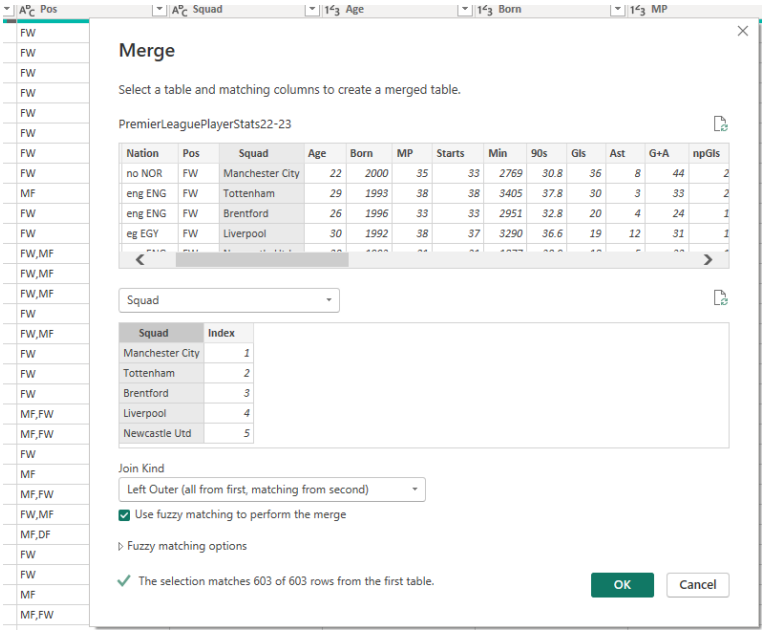


Figure 33: Merging the Queries on The Squad Value

3.6.7 Add Identifier to Fact Table And Untick Bottom Selection Box

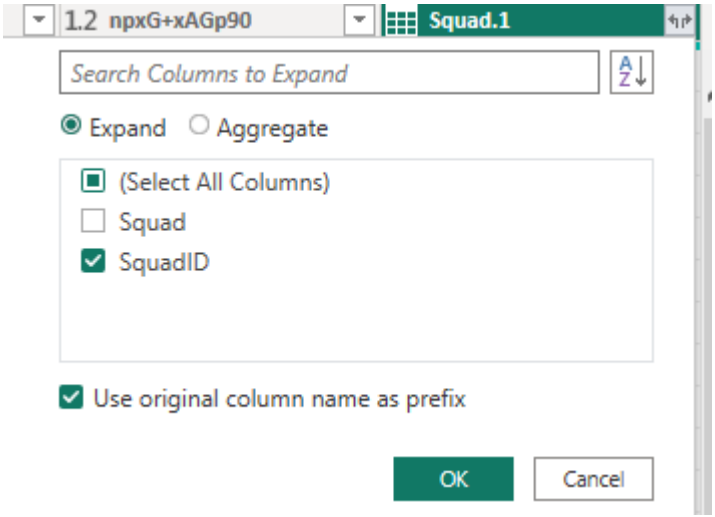


Figure 34: Selecting the SquadID as an Identifier for table

3.6.8 View Results

1.2 xGp90	1.2 xAGp90	1.2 xG+xAGp90	1.2 npxGp90	1.2 npxG+xAGp90	1²3 SquadID
1.2	0.92	0.17	1.1	0.75	0.92	1
0.74	0.57	0.18	0.75	0.44	0.63	2
0.55	0.57	0.15	0.73	0.4	0.56	3
0.79	0.59	0.21	0.8	0.51	0.72	4
0.96	0.84	0.13	0.97	0.72	0.85	5
0.69	0.48	0.11	0.59	0.48	0.59	6
0.65	0.3	0.29	0.59	0.3	0.59	7
0.58	0.47	0.11	0.58	0.43	0.54	8
0.63	0.29	0.23	0.52	0.29	0.52	7
0.49	0.67	0.1	0.77	0.39	0.49	9
0.65	0.32	0.24	0.56	0.26	0.5	7
0.46	0.3	0.1	0.4	0.3	0.4	10
0.6	0.38	0.08	0.46	0.34	0.42	11
0.6	0.38	0.08	0.46	0.34	0.42	11
0.47	0.29	0.11	0.4	0.29	0.4	5
1.11	0.41	0.19	0.59	0.41	0.59	4
0.83	0.29	0.23	0.52	0.29	0.52	1
0.7	0.61	0.17	0.78	0.58	0.75	7
0.71	0.42	0.07	0.49	0.42	0.49	12
0.44	0.25	0.15	0.39	0.22	0.37	13

Figure 35: New Row with SquadID Added

3.6.9 Remove All Corresponding Data From Fact Table

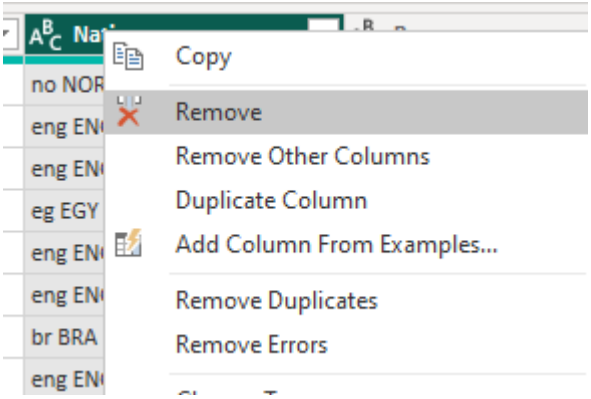


Figure 36: Removing All Data That Is No Longer Needed

3.7 Removing Columns

When creating the stats table, there was some overlap of statistics from the other tables. This meant that duplicate columns were presents. I decided to remove all the duplicate columns by right clicking and clicking remove in the power query editor.

3.8 Calculated Columns

3.8.1 Age Group

The below code created a table that assigned a column in the “PlayerData” table with either “Rising Talent”, “Emerging”, “Peak” or “Declining” This was used to help categorise our players.

```
AgeGroup =
    SWITCH(
        TRUE(),
        'PlayerData'[Age] >= 15 && 'PlayerData'[Age] <= 18, "15-18 / Rising Talent",
        'PlayerData'[Age] >= 19 && 'PlayerData'[Age] <= 23, "19-23 / Emerging",
        'PlayerData'[Age] >= 24 && 'PlayerData'[Age] <= 31, "24 - 31 / Peak",
```

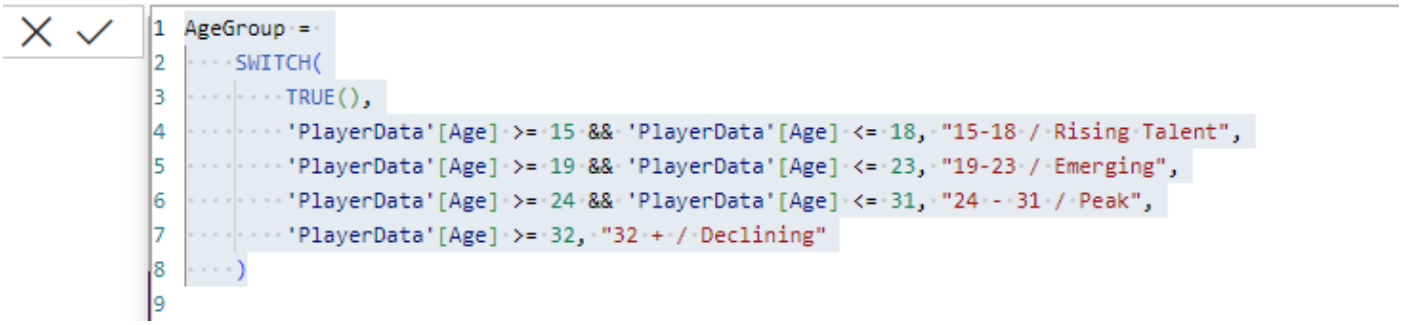


Figure 37: New Column Created Based on Existing Data

3.9 Dax Measures

3.9.1 Average Age

This calculation uses the average built in function to work out the average age of players.

```
1 AvgAge = AVERAGE(PlayerData[Age])
```

Figure 38: Average Age Formula

3.9.2 Arrows

These calculations aid the upwards and downwards arrow showcased on some of the pages. If a threshold is above or below a specific value, then the arrow changes direction.

```
1  Icon =
2  VAR PositiveIcon = UNICHAR(9650)
3  VAR NegativeIcon = UNICHAR(9660)
4  VAR PlayerCount = COUNTROWS(DISTINCT(PlayerData[PlayerID]))
5  VAR Result =
6      IF(PlayerCount > 30, NegativeIcon,
7          IF(PlayerCount >= 24 && PlayerCount <= 28, PositiveIcon,
8              NegativeIcon)
9      )
10 RETURN Result
11
```

```
1 Icon 1 =
2 Var PositiveIcon = UNICHAR(9650)
3 Var NegativeIcon = UNICHAR(9660)
4 Var Result = IF([Total Goals] > 45,
5 PositiveIcon, NegativeIcon)
6 RETURN Result
```

```
1  Icon2 =
2  VAR PositiveIcon = UNICHAR(9650)
3  VAR NegativeIcon = UNICHAR(9660)
4  VAR Result =
5      IF(
6          AVERAGE(StatsData[Cmp%]) > 70.1,
7          PositiveIcon,
8          NegativeIcon
9      )
10 RETURN Result
11
```

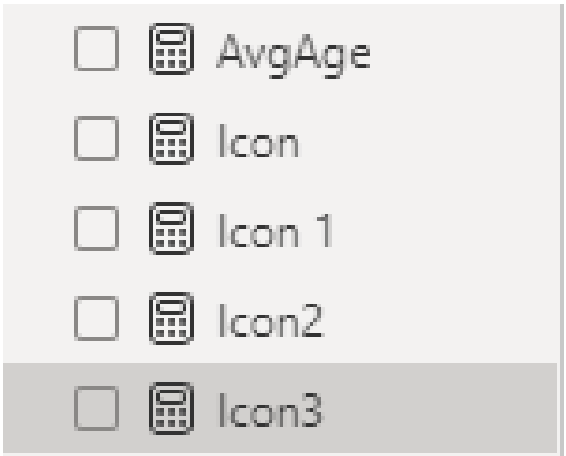


Figure 39: Arrow Measures

```
1  Icon3 =
2  VAR PositiveIcon = UNICHAR(9650)
3  VAR NegativeIcon = UNICHAR(9660)
4  VAR Result =
5      IF(
6          SUM(StatsData[TotDist]) > 61000,
7          PositiveIcon,
8          NegativeIcon
9      )
```

Figure 40: Code for Arrows

3.9.3 Over Performance xG/Under Performance xG

This measure creates a table which works out how many goals a player scored more than their expected goals number as well as how many goals they underperformed their xG.

```
1 OverPerformanceXG = SUMX('StatsData', 'StatsData'[Gls] - 'StatsData'[xG])
```

Figure 41: DAX For Underperforming Player xG

```
1 UnderPerformanceXG = SUMX('StatsData', 'StatsData'[Xg] - 'StatsData'[Gls])
```

Figure 42: DAX For Underperforming Player xG

3.9.4 Unique Nationalities

This piece of Dax counts the total nationalities. This was used in the Club Analysis page where the diversity of each team is documented.

```
1 UniqueNationalities = DISTINCTCOUNT('Nation'[Nation])
```

Figure 43: Dax for Unique Nationalities

3.9.5 Total Goals

```
1 Total Goals = SUM('StatsData'[Gls])
2
```

Figure 44: Dax for Total Goals Scored

4.0 Star Schema Facts and Dimensions

Below is the completed Star Schema which uses the middle fact table as a joiner with primary keys from other tables as foreign key. All relationships are accurate and were manually created. To create a relationship in PowerBI you simply drag the primary key of one table to another (foreign key) and PowerBI and set the "Cardinality" to the type of relationship you wish to implement. It is possible to change the relationship type but, in this example, PowerBI managed to crate the correct relationships.

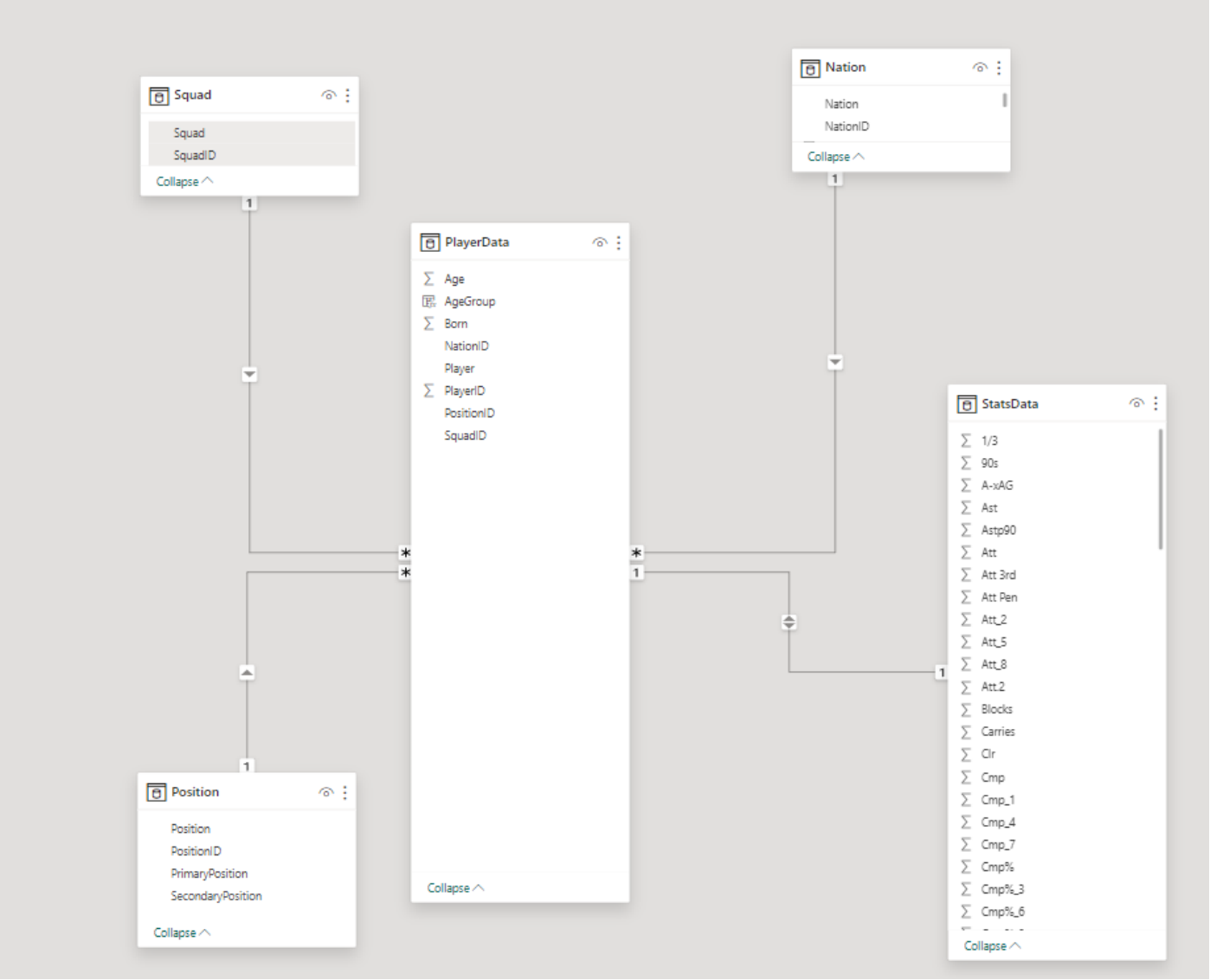


Figure 45: Created Star Schema

5.0 Findings Based On Analysis

5.1 KPI Data Visuals

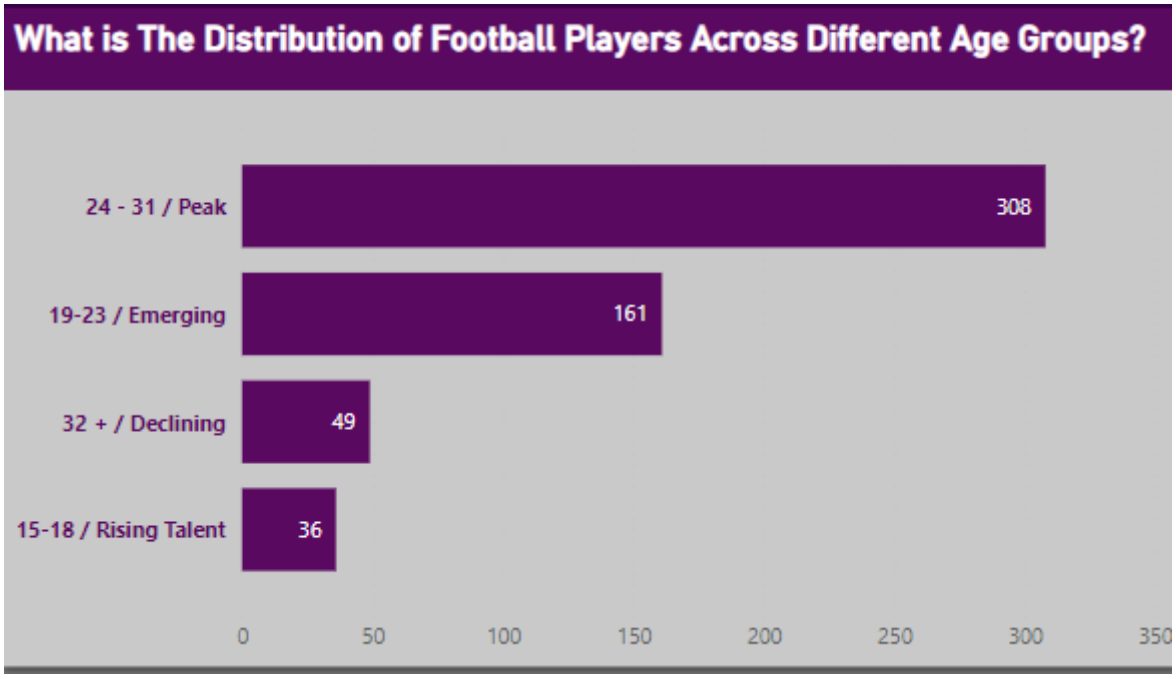


Figure 46: Footballing Age Distribution

The above diagram visualises the age distribution of players within the premier league. This table uses the previously created column which is shown on the Y axis. The findings make for interesting reading showing that a large majority of players are in the peak years of their footballing career thus strengthening the previous statement of the league attracting players in their peak years. This chart visual was chosen due to it clean aesthetic in combination with the data returning some small fields which may have led to harder viewing in alternatives like a pie chart.

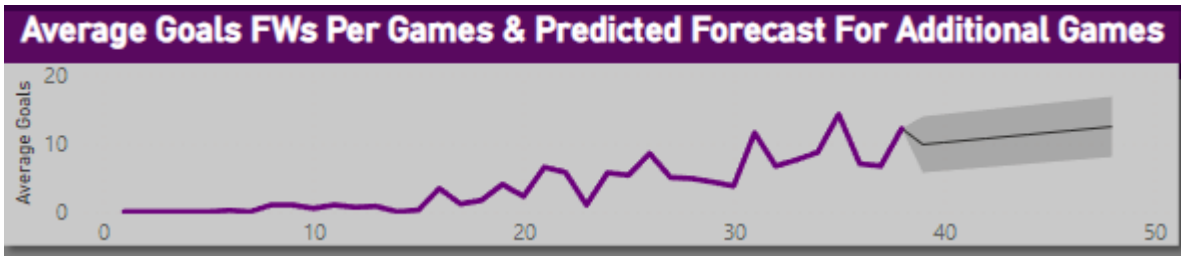


Figure 47: Average Goals Per Game

This graphic makes for interesting viewing as it calculates the average goals for forwards based on the number of games they play. Additionally, it uses the forecast feature which based on the data predicts the number of goals that would be scored over additional games which are not included within the dataset. The use of a line chart is crucial as it easily shows trends in games and provides an overall upward trajectory of the more games, the more goals. It is also not possible to apply a forecast to many other graph types.

Figure 8 Map of Diversity for Squads

Figure 48: Team Squad Diversity Visual

This visual is when a user of the dashboard selects one of the twenty teams to get an analysis for. It counts the number of nationalities that are present within the first team squad and plots it on a map chart. The map chart was a clear decision as it easily visualises diversity and the size of the bubble is based on how many players are from each nation. As we can see from above, the greatest pool of players comes from England.

5.2 BI-Dashboard

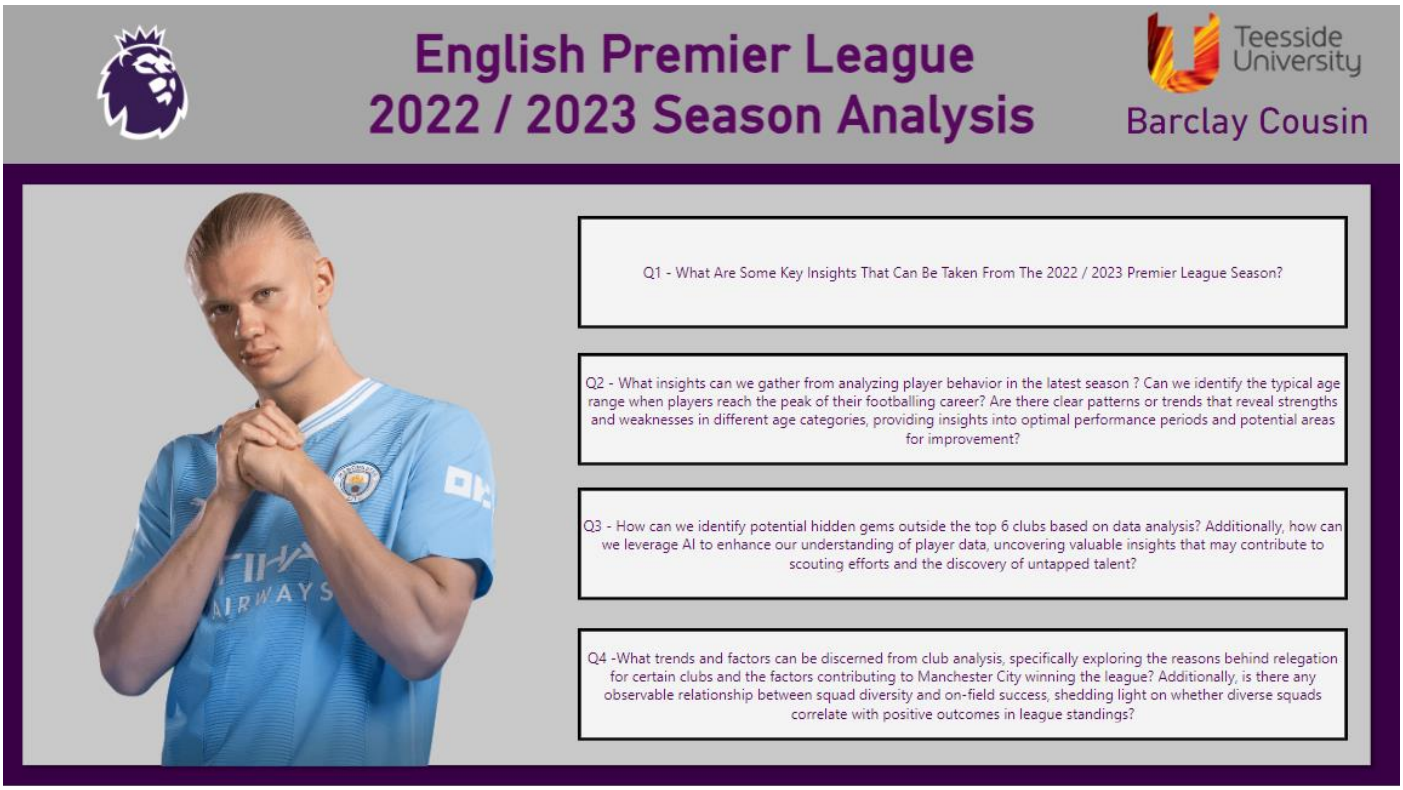


Figure 49: Homepage

The Dashboard presents a visually aesthetic and descriptive homepage offering comprehensive insights into the posed questions to be answered within the report. Use of colour is present throughout the entirety of the report and the image of the division's top goal scorer is well placed to add to the aesthetics.

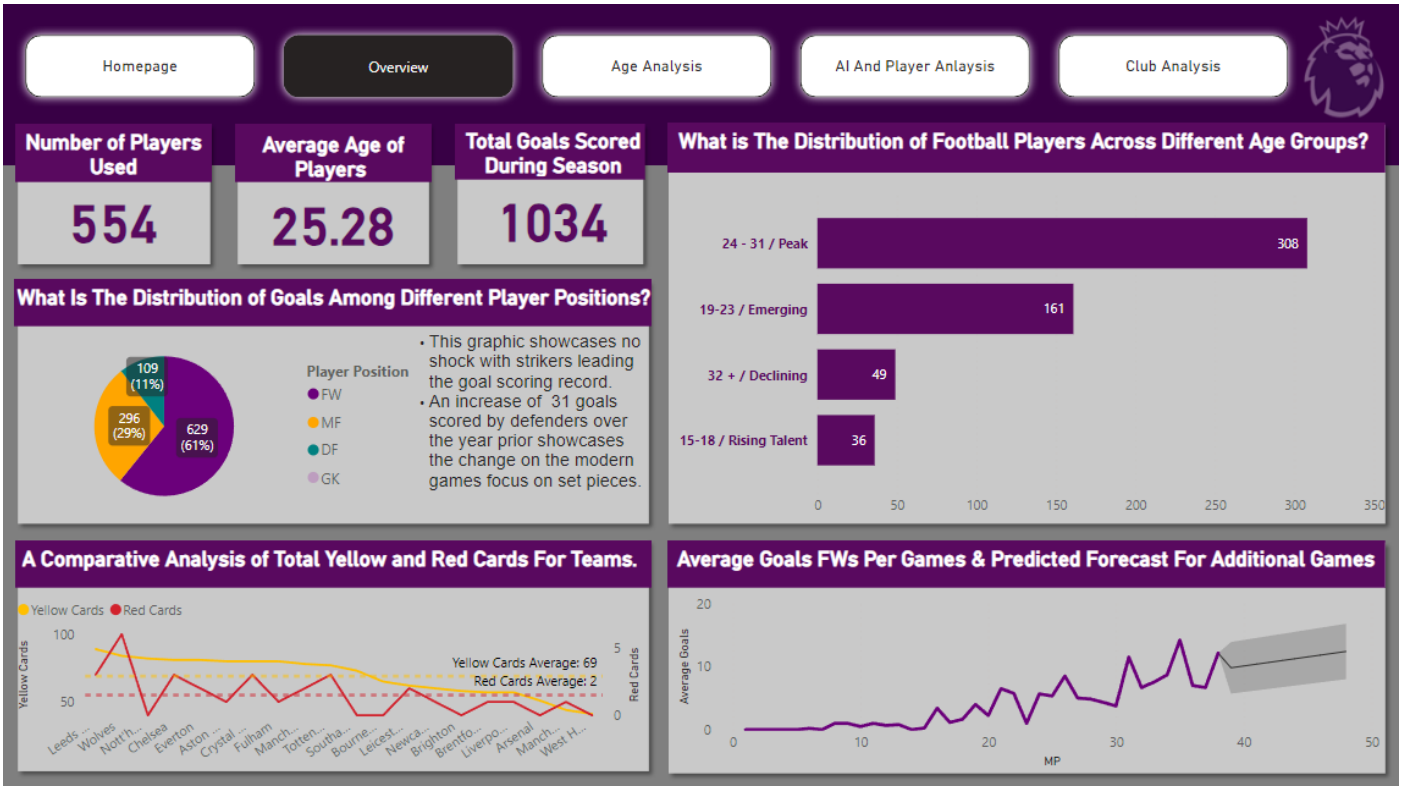


Figure 50: Overview Page

Above is the overview page of the report. This page visualises some key data that will be of interest to a general user. Three cards are used to demonstrate some KPIs that a reader will find useful. A bar chart has been used to portray the ages of players within the division showing most players are aged 24-31. Finally, there are two-line charts used. One showcases both the yellow and red cards given to teams within the division and the other for goals with a specific focus on the forecast feature.

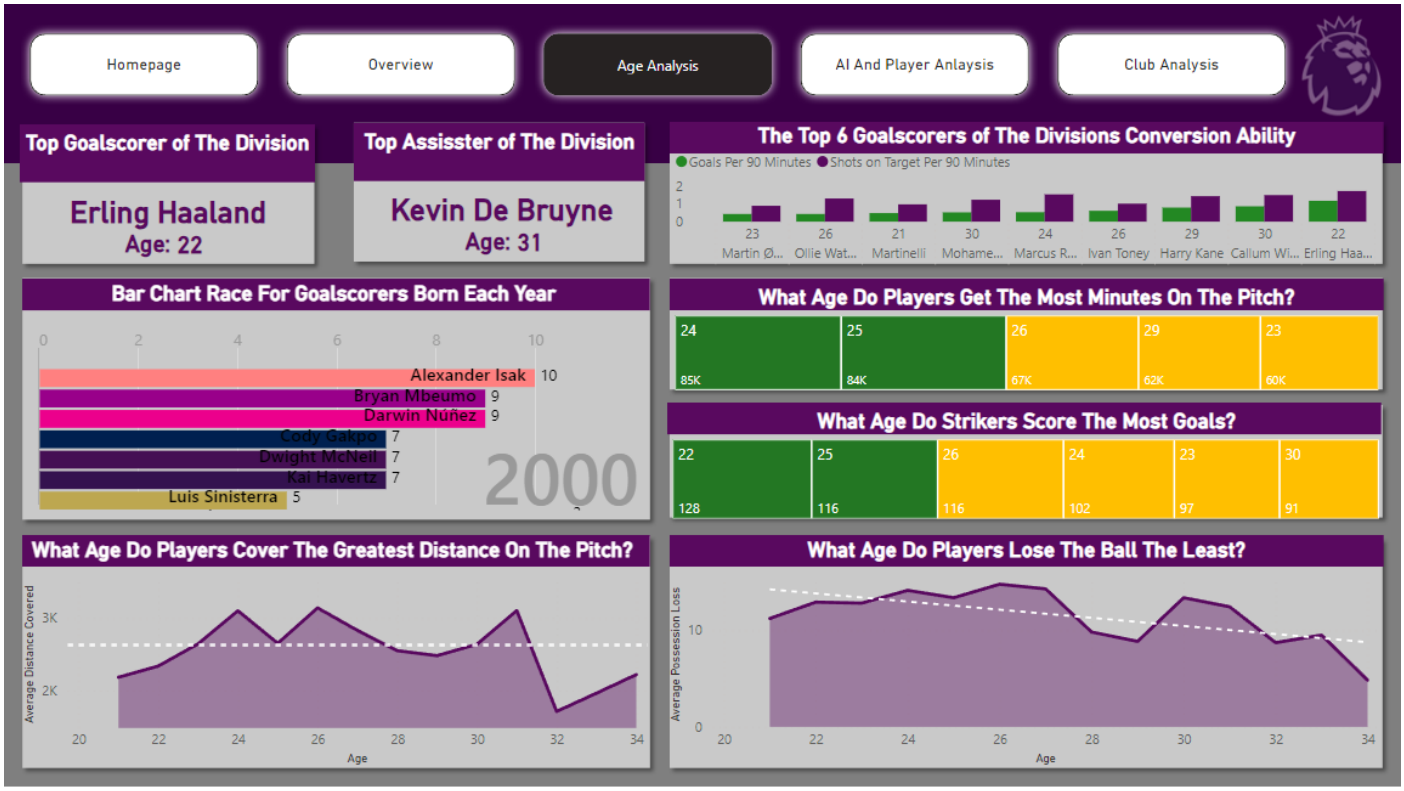


Figure 51: Age Analysis Page

This page answers questions relating to when a player is in their prime footballing years. A varied number of charts were used to help answer this question from cards for static data, conversion ability for strikers with their age, and line charts to understand when a player is their most physical and experienced enough to lose the ball the least. An animated chart was used that visualises the goal scorers there were born in each year with the total goals they scored. The bar chart race will be animated upon viewing the Pbix file.

Figure 52: AI And Player Analysis

This chart answers questions that may be posed by scouting departments to get some key statistics on their players and new signings. Two AI Graphics have been used which can help understand the goals scored within the league with artificial intelligence. The other visuals showcase players who both underperformed and overperformed when it came to goalscoring. Finally, two bar charts have been used to highlight some stand-out players outside the top 6 which scouting departments should look at further, as it can be a great insight into future signings for their team for their team.

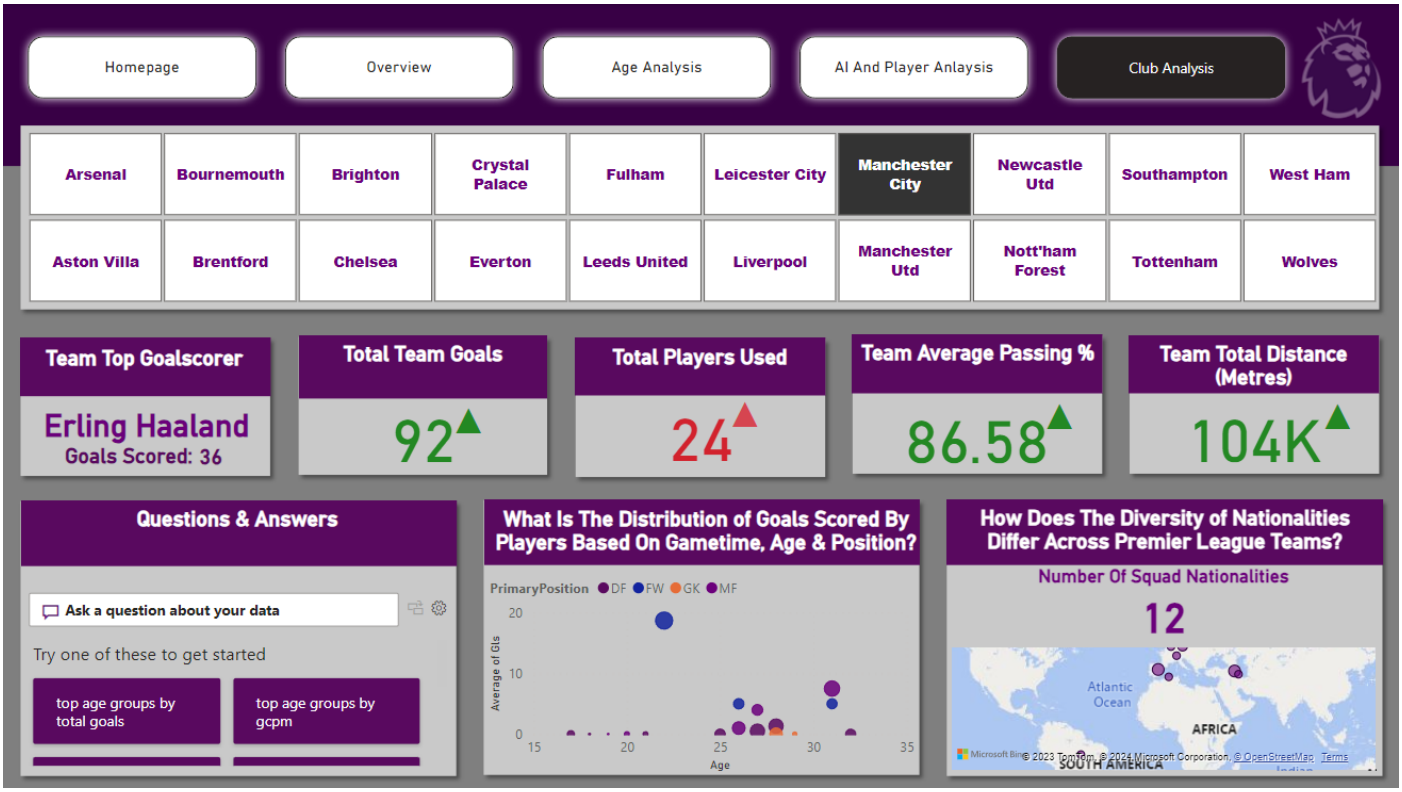


Figure 53: Club Analysis

The above page provides specific insights into individual clubs. The use of red, amber, and green matrix has been implemented to identify how it ranks against the average. Specific insights can be taken such as the reasons Manchester City won the league. It was clear to see that they were a hardworking team that scored a lot of goals. The Q & A feature was also used on this page which is powered by artificial intelligence. On the opposite spectrum for a team like Leeds United who got relegated, there would be a lot of red for the KPIs.

5.3 Findings

- Wolves need to work on their discipline with them amounting the most red cards in the league (6) and second most yellow cards in the league (84) which are both well above the average of the division.
- Strikers hit their peak in their mid-20s and the scouting department for teams should acknowledge this and aim to sign attackers between the ages of 23 and 27 for optimal goalscoring output.
- Players are at their physical (average distance covered on the pitch) peak between the ages of 24 and 28 and scouting departments should set a specific focus for midfielders between this age.
- Some key stand out performers in midfield are James Ward-Prowse, Moises Caicedo and Declan Rice and should be looked into by top clubs as potential signings.
- Despite Leeds United's relegation their goal threat Rodrigo should be looked at by a top division club due to him outscoring his xG by 4.9 thus meaning with his limited service he manged to convert his chances.
- Chelsea's lacklustre season can be explained by their striker Kai Havertz topping the xG underperformance table with 4.6 meaning a new clinical striker should be signed.
- Brazilian players scored almost 10% of goals in the premier league showing that they are still a powerhouse for producing footballing talent.
- Manchester City have one of the least diverse squads in the division with 12 nationalities on record which could potentially show having a tight knit squad with lack of diversity could be a notable direction for clubs to take.
- Southampton had the highest number of players used in the season (36) and the lack of player cohesion could be a deciding factor of their rock bottom finish.

5.4 Recommendations & Conclusion

To summarise, the findings showcase a player discipline issue for Wolves which must be addressed, key metrics relating to the age bracket for midfielder and attacking players was concluded and noteworthy midfielders and defenders were highlighted outside of the top 6 clubs for prospects. Despite Rodrigo playing for relegated Leeds United, he showcased goal-scoring proficiency of the highest standard. Chelsea's challenges have a direct correlation Kai Havertz underperformance highlight the necessity for bringing a talisman to the club. The lasting impact of Brazilian players to the premier league with consideration of squad diversity instil the intricate nature of the English topflight. Southampton's struggles underscore the importance of team cohesion for sustained success within the league with them using far too many players within the campaign. These are just some of the key points which can be addressed to enable clubs to have long term success in the future.

References

Kaggle. (2022). Premier League Player Stats 2022-23. Kaggle. Available at: https://www.kaggle.com/datasets/ameyaranade/premier-league-player-stats-2022-23/data (Accessed: 6 January 2024).

Microsoft Community. (2017) 'Power BI Map 3-Letter Country Code', Available at: https://community.fabric.microsoft.com/t5/Desktop/power-bi-map-3-letter-country-code/m-p/171395 (Accessed: 1 January 2024).